

Annotations and the Digital Humanities Research Cycle: Implications for Personal Information Management

Marie-Eve Bélanger

Faculty of Information, University of Toronto

140 St. George Street

Toronto, Ontario M5S 3G6

+1 (647) 520-0675

me.belanger@utoronto.ca

ABSTRACT

The proposed study assesses the creation, use and organization of annotations in the digital humanities research cycle. It is argued that the gap between digital and physical reading practices creates complex personal information collections, forcing the scholar to cope with information fragmentation by adapting his practices within the constraints of the research process. A poster is proposed outlining a research design and early findings regarding this issue.

Keywords

Annotation, personal information management, scholarly research model

1. INTRODUCTION

Increasingly, primary and secondary textual scholarly resources are available to the humanities scholar in digital form. Such access affords new opportunities for retrieving, exchanging, and storing documents. Recent developments in the digital humanities community have yielded multiple software and web applications supporting the interpretative process of scholars, offering tools for textual and linguistic analysis. However, a large number of academics still find it necessary to print out digital documents to create freeform, idiosyncratic annotations and to profit from the spatiality afforded by the materiality of paper. The available tools and new technologies, as well as the scholar's habits, inevitably create a hybrid information space where information items may be available in multiple formats, organized according to numerous classification schemes, and accessed in various ways. The scholar may then be forced to cope with this information fragmentation by adapting his practices within the constraints of the research process.

In this poster I lay out my reasoning and progress towards understanding the humanities scholar's evolving personal information management (PIM) practices throughout the research cycle. I argue that renewed attention to the diversity of scholarly activity and related annotation practices can inform current research in personal information management.

2. PIM AND THE SCHOLARLY PROCESS IN THE HUMANITIES

PIM refers to the activities performed by an individual in order

to “acquire or create, store, organize, maintain, retrieve, use and distribute the information needed to complete tasks” [6]. To accomplish these activities, individuals manipulate information items, physical or digital representations of the information. This study assesses how scholars create, use and organize the information items related to their research project, from ideation to dissemination, explicitly addressing the role of annotations as pervasive information items. This constantly evolving cluster of related information items may be referred to as a personal information collection. While researchers currently disagree on how to correctly define a “collection” [6], this poster aims to provide functional dimensions of personal information collections, hereby adding to the ongoing dialogue in the PIM research community.

The researcher's information collection may change in size and nature throughout the research cycle, according to various phases. While the notion of the humanities scholar as a solitary researcher who values browsing and footnote chaining is now widely understood [3, 14], researchers are still struggling to define the basic, primordial phases of the research process [12, 13, 5, 2, 11, 4]. However, while the terminology, span, and breadth of the research phases vary from model to model, they each touch on common and discrete activities integral to scholarly work across domains, termed “scholarly primitives” by Unsworth [15]. These activities, such as discovering, comparing, referring and sampling, are the basic, constant and recursive modes of interactions between the researcher and his research material.

While these models of scholarly research shed some light on the activities of the scholar and their inherent processes, they fail to account for the variety of annotation practices permeating the various stages of the research cycle. Indeed, most of these models consider the creation and organization of annotation to be strictly relevant to the preparation and elaboration phases of the cycle. However, the typology of reading devised by O'Hara [8] reveals that 1) different types of reading occur throughout the research cycle as seen in Table 01 and, 2) different annotation practices are associated with these reading goals. This is echoed by Palmer and Cragin [11] who hint at the pervasiveness of annotation practices within the scholarship cycle. Building on Unsworth's notion of the scholarly primitive, Palmer and Cragin defined finer grained micro-processes as “information work primitives”. Annotation, a type of information work primitive, functions as an articulation device, aligning the different levels, activities and indeed, the different phases of work necessary to the completion of a project. The

Table 1. Comparison of research cycle models and of reading goals using Chu’s Research-Phases model as a basis.

Chu (1999)	Stone (1980)	Bradley (2008)	O’Hara (1996)
1. Idea stage	1. Thinking and talking to people		• Reading to learn
2. Preparation Stage	2. Reading what has already been done in the field 3. Studying original sources and making notes	1. Reading and annotating	• Reading for research • Reading to summarize
3. Elaboration Stage			• Reading for critical review
4. Analysis and writing stage	4. Drafting write-up 5. Revising the final draft	2. Developing interpretation	• Reading while writing from multiple sources • Reading to search / reading to answer questions • Proof-reading • Reading for text revision
5. Dissemination Stage		3. Presentation of interpretation	
6. Further writing and dissemination stage			

variety of reading types and consequently of annotation practices may explain why, despite having access to software supporting annotation, academics still tend to print and physically mark up digital documents, rapidly increasing the amount of information items to manage over the course of a project.

3. A FUNCTIONAL TYPOLOGY OF ANNOTATION

As noted above, the personal information collection of the scholar grows both in size and complexity due to the hybridity of the research activities. A functional typology, based on the varying dimensions and formats of annotations may help explain the ongoing gap between digital and paper forms of reading practices and thus characterize the creation, use and organization of information items by the humanities scholar.

Annotations, understood as information items created and manipulated by the researcher, may be used differently according to their content. Several PIM studies, as summarized by Barreau [1] have divided information items according to three categories: archived (long-term value, but unrelated to current work), working (frequently used) and ephemeral (short lifespan and used for non-routine tasks). Boardman and Sasse [1] further refined this classification by proposing four categories: active (working and ephemeral information), dormant (potentially useful but inactive information), not useful and not assessed. This latter categorization may provide valuable insight into the diversity of annotation practices throughout the research process when coupled with the reading goals devised by O’Hara [8] as seen in Table 2. Furthermore, recent research has revealed multiple types of primary uses for annotation: to remember, to think, to clarify, and to share [10]. This poster argues that some of these high-level purposes may be more prevalent than others at various stages of the research cycle.

Types of use as well as organization strategies may also be influenced by the information form of annotations. The choice of

tools and medium seems to be related to the formal/informal and explicit/tacit dimensions of annotations, as described by Marshall [7]. Freeform annotations are generally informal and incomplete, demanding to be made quickly and in a minimally disruptive manner. From a cognitive standpoint, paper is the support of choice for readers of printed documents to quickly offload their working memory by creating unself-conscious, informal and incomplete annotations, thus avoiding the loss of information due to an overly disruptive process [9]. In addition, paper may more readily support idiosyncratic annotation methods, such as special signs or individual correction marks, as well as more graphic markings such as margin bars, circling and underlining. While a vast amount of systems developed in the human-computer interaction community may readily support these informal markings by using tablet computers, these technologies have yet to be widely adopted by the digital humanities community, accentuating the fragmentation of information.

4. CONCLUSION

Printing digital documents for reading and annotating purposes leaves scholars with an imbalanced, dual-medium representation. This may increase the complexity of the scholar’s personal information collection in 2 distinct ways. The digital and physical copies, unaltered by annotations, represent two separate information items, differentiated by form. These two items are independent from each other, have their own life cycles (one item may be discarded before the other) and are possibly dependent on different organization schemes. Additionally, the content of these items can be independently modified: the printed document, once replete with notes, does not correspond structurally or semantically to its digital version. This poster is centered on the issues emerging from this gap and their effects on personal information management practices of humanities scholars. Early findings, based on a review of the relevant literature as well as interviews and observation sessions of humanities scholars, are reported in the poster.

Table 2. Types of information items created according to reading goals and research cycle phases.

		Phases of the Research Cycle (Chu)				
		Idea	Preparation	Elaboration	Analysis/Writing	Dissemination
Annotations made while...	Reading to learn	ephemeral	dormant	dormant	dormant	dormant
	Reading for research	-	working	working	working	dormant
	Reading to summarize	-	working	working	working	dormant
	Reading for critical review	-	-	working	working	dormant
	Reading while writing from multiple sources	-	-	-	working	dormant
	Reading to search/reading to answer questions	-	-	-	ephemeral	-
	Proof-reading	-	-	-	ephemeral	-
	Reading for text revision	-	-	-	ephemeral	-

5. ACKNOWLEDGMENTS

The author expresses her gratitude to Dr. Matt Ratto and Dr. Alan Galey for their continued support and constructive comments throughout this research.

6. REFERENCES

- [1] Barreau, D. 2008. The persistence of behavior and form in the organization of personal space. *Journal of the American Society for Information Science and Technology*, 59, 2, 307-317
- [2] Bradley, J. 2008. Thinking about interpretation: Pliny and scholarship in the humanities. *Literary and Linguistic Computing*, 23, 3, 263-279.
- [3] Brockman, W. S., Neumann, L., Palmer, C. L., & Tidline, T. J. 2001. *Scholarly Work in the Humanities and the Evolving Information Environment*. Washington, DC: Digital Library Federation and the Council on Library and Information Resources.
- [4] Case, D. O. 1986. Collection and organization of written information by social scientists and humanists: a review and exploratory study. *Journal of Information Science*, 12, 3, 97-104.
- [5] Chu, C. M. 1999. Literary critics at work and their information needs: a research-phases model. *Library & Information Science Research*, 21, 2, 247-273.
- [6] Jones, W. 2007. Personal Information Management. *Annual Review of Information Science and Technology*, 41, 1, 453-504.
- [7] Marshall, C. C. 1998. Toward an ecology of hypertext annotation. in *Proceedings of ACM Hypertext '98*, Pittsburgh, PA (June 20-24, 1998), 40-49.
- [8] O'Hara, K. 1996. *Towards a Typology of Reading Goals. Technical Report EPC-1996-107*. Cambridge: Rank Xerox Research Center.
- [9] O'Hara, K. P., Taylor, A., Newman, W., & Sellen, A. J. 2002. Understanding the materiality of writing from multiple sources. *International Journal of Human-Computer Studies*, 56, 3, 269-305.
- [10] Ovsiannikov, I. A., Arbib, M. A. & Mcneill, T. H. 1999. Annotation technology. *International Journal of Human-Computer Studies*, 50, 4, 329-362.
- [11] Palmer, C. L. & Cragin, M. H. 2008. Scholarship and disciplinary practices. *Annual Review of Information Science and Technology*, 42, 1, chapter 5.
- [12] Stone, S. 1980. CRUS humanities research programme. In *Humanities information research: Proceedings of a seminar; Sheffield 1980BLR&DD Report No. 5588*, Centre for Research on User Studies, University of Sheffield, Sheffield, England, 15-26.
- [13] Stone, S. 1982. Humanities scholars: information needs and uses. *Journal of Documentation*, 38, 4, 292-313.
- [14] Toms, E., & O'Brien, H.L. 2008. Understanding the information and communication technology needs of the e-humanist. *Journal of Documentation*, 64, 1, 102-130.
- [15] Unsworth, J. 2000. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this? In *Symposium on Humanities Computing: Formal Methods, Experimental Practice*. King's College, London. Retrieved November 1, 2009, from <http://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html>